

# 3did: interacting protein domains of known three-dimensional structure

Amelie Stein<sup>1</sup>, Robert B. Russell<sup>1,2</sup> and Patrick Aloy<sup>1,\*</sup>

<sup>1</sup>EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany and <sup>2</sup>EMBL, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton Hall, Cambridge CB10 1SD, UK

Received July 22, 2004; Revised and Accepted September 23, 2004

## ABSTRACT

**The database of 3D Interacting Domains (3did) is a collection of domain–domain interactions in proteins for which high-resolution three-dimensional structures are known. 3did exploits structural information to provide critical molecular details necessary for understanding how interactions occur. It also offers an overview of how similar in structure are interactions between different members of the same protein family. The database also contains Gene Ontology-based functional annotations and interactions between yeast proteins from large-scale interaction discovery studies. A web-based tool to query 3did is available at <http://3did.embl.de>.**

## INTRODUCTION

Proteins are social molecules and most biological processes require many of them to interact. This has encouraged many projects aimed at finding protein functions based on the detection of their relationships. Genome-scale interaction discovery approaches, such as the two-hybrid system (1–5) and affinity purification (6,7) have suggested thousands of protein–protein interactions. *In silico* approaches have also predicted many interactions with levels of accuracy similar to those determined experimentally (8). Put together, all these interactions have uncovered many aspects of protein connectivity but without critical molecular details often necessary to understand their function. Another difficulty is that it is often impossible to distinguish between direct physical interactions and functional associations that may not involve direct atomic contacts between macromolecules. Currently, atomic details of interactions are present in high-resolution three-dimensional (3D) structures of protein complexes but this information is scarce and has been largely overlooked in large-scale studies. The database of interacting domains of known 3D structure (3did) exploits structural information to provide atomic details for thousands of direct physical interactions between proteins.

## 3did CONTENT

Proteins are composed of modular elements (domains) that to a great extent determine their structure, function and interaction partners. We thus decided to structure our database on domains rather than full-length proteins. 3did obtains the high-resolution structures of individual proteins and complexes from the Protein Data Bank (PDB) (9). Pfam (10) domains are then assigned to each individual protein and interactions between them are computed and the information stored.

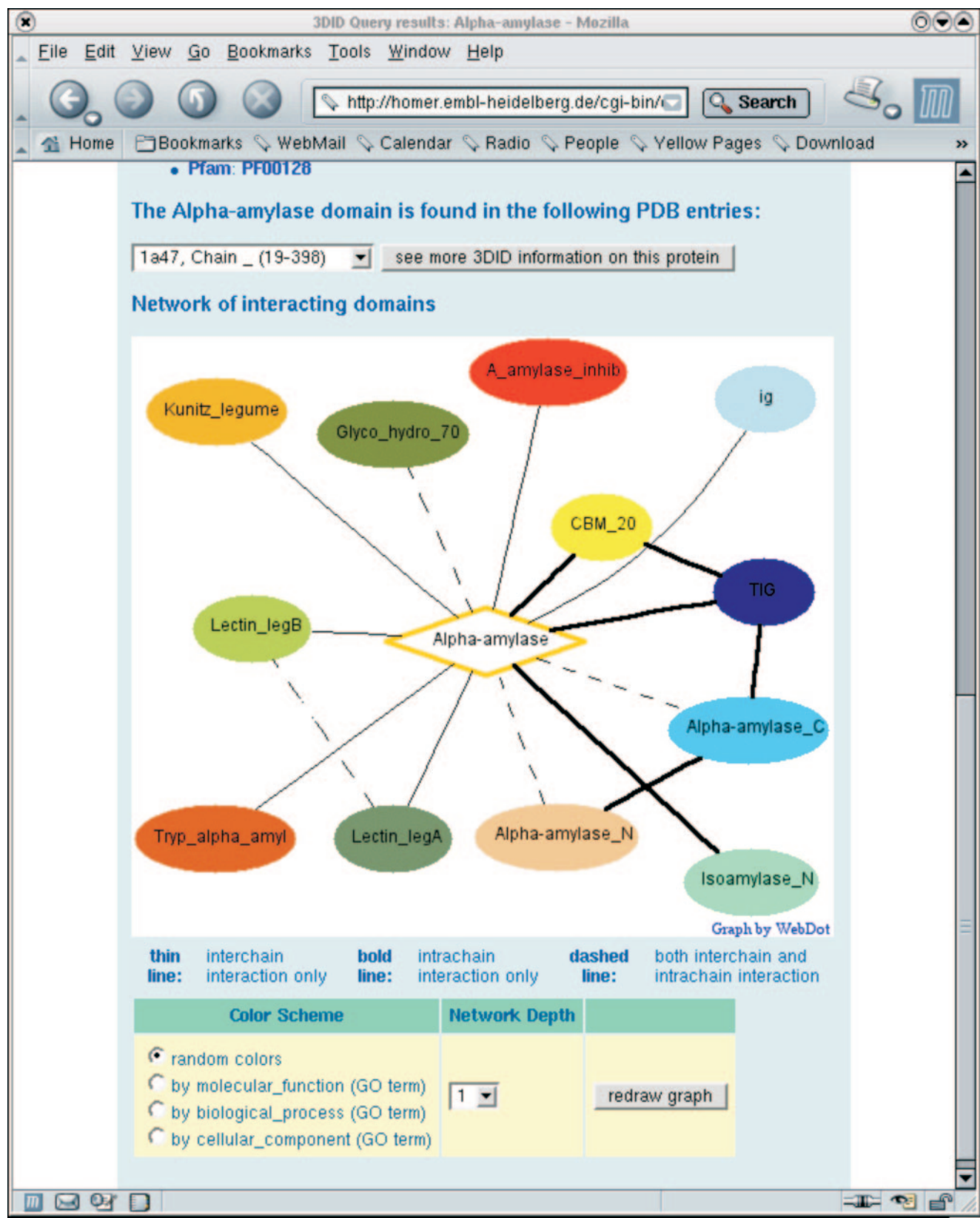
Currently, 3did includes information on 50 700 protein chains of known 3D structure making a total of 48 426 domain–domain interactions. Of these, 13 482 occur between domains in the same chain (i.e. intra-molecular) and 34 944 between domains lying in different proteins (i.e. inter-molecular). We grouped these interactions into 2535 types according to the Pfam domains mediating them. Of these 411 always interact within the same polypeptide chain (intra-molecular), 1765 are only seen in different chains (inter-molecular), and 359 containing both intra- and inter-molecular interactions. When available, 3did also contains functional information about the interacting domains. Gene Ontology (GO) (11) terms for molecular function, biological process and cellular component could be assigned to 1325, 1122 and 480 families, respectively. The database also contains 1128 links between known structures and interactions between yeast proteins determined experimentally as defined in MIPS (12). New 3D structures are incorporated weekly and major updates take place whenever a new version of Pfam is released. Up-to-date statistics on 3did contents can be found in the website.

## 3did USAGE AND FEATURES

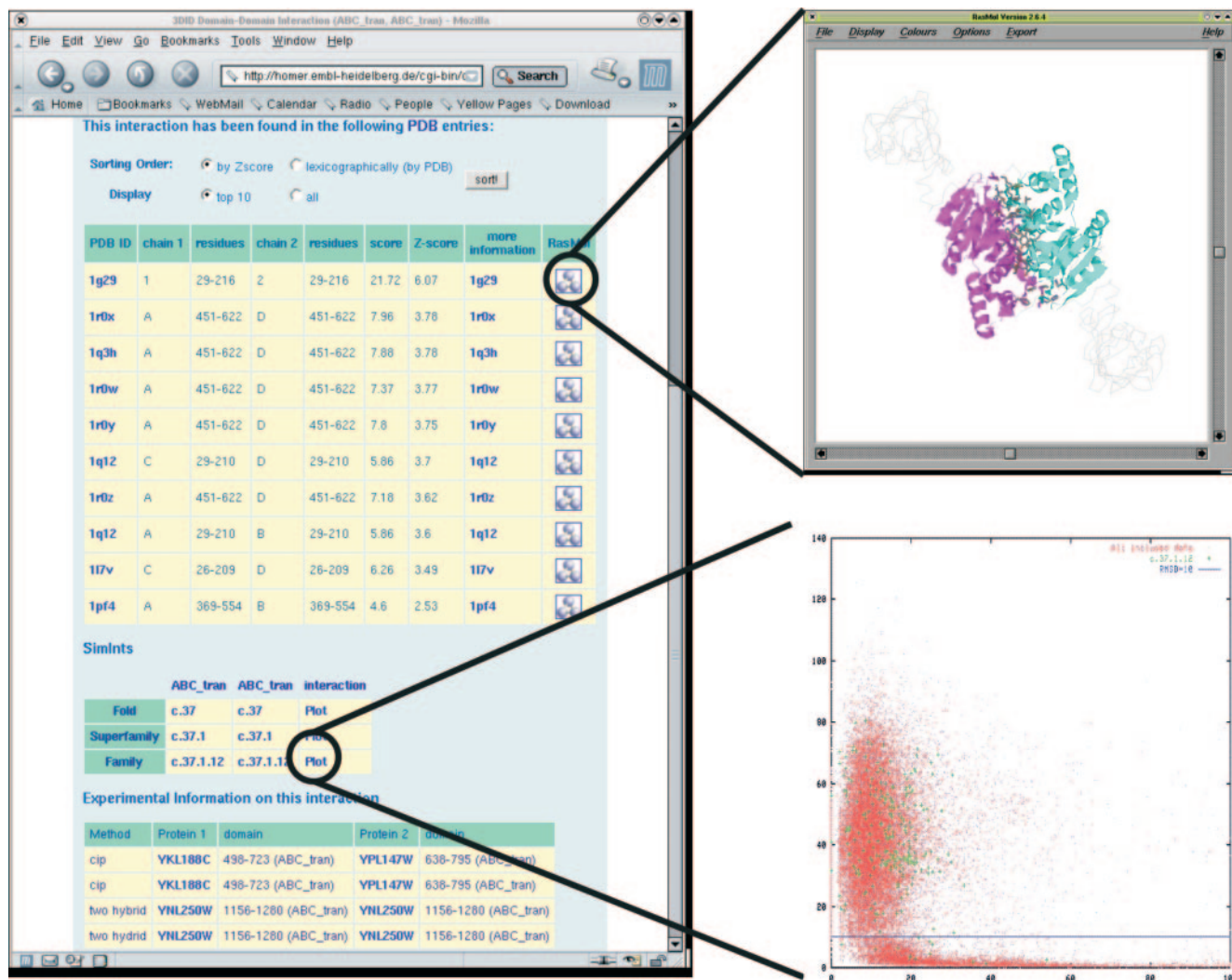
The standard way of accessing the database is by querying it with a particular domain. When doing so, 3did will show all domains that physically interact with our domain of interest and for which the 3D structure of the interaction is known. We computed physical interactions by requiring at least

\*To whom correspondence should be addressed. Tel: +49 6221 387 305; Fax: +49 6221 387 517; Email: [aloy@embl.de](mailto:aloy@embl.de)

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact [journals.permissions@oupjournals.org](mailto:journals.permissions@oupjournals.org).



**Figure 1.** Network of interacting domains. The domain of interest (e.g. Alpha-amylase) is shown as a rhombus and the interacting domains as ellipses. By default, the interaction graph is colored randomly, but it can be colored based on functional GO terms. Bold lines represent intra-chain interactions (i.e. between domains in the same protein), thin lines indicate inter-chain interactions (i.e. between domains in different proteins) and dashed lines show interactions where both types have been seen. The network depth can be changed to include indirect interactions.



**Figure 2.** Example of the information given for two interacting domains. The table shows a subset of the structures where this interaction is present, with information on the location of the interacting domains (PDB numbering) as well as the interaction scores as described previously (13). By clicking at the *rasmol* icon, the user will get a 3D display of the complex with the two interacting domains shown in *ribbons* and the contacting residues in *ball-and-stick* (magnification top). Where possible, there are also links to SimInt showing how conserved is the interaction type (magnification bottom) and to yeast proteins found experimentally to interact that contain these domains.

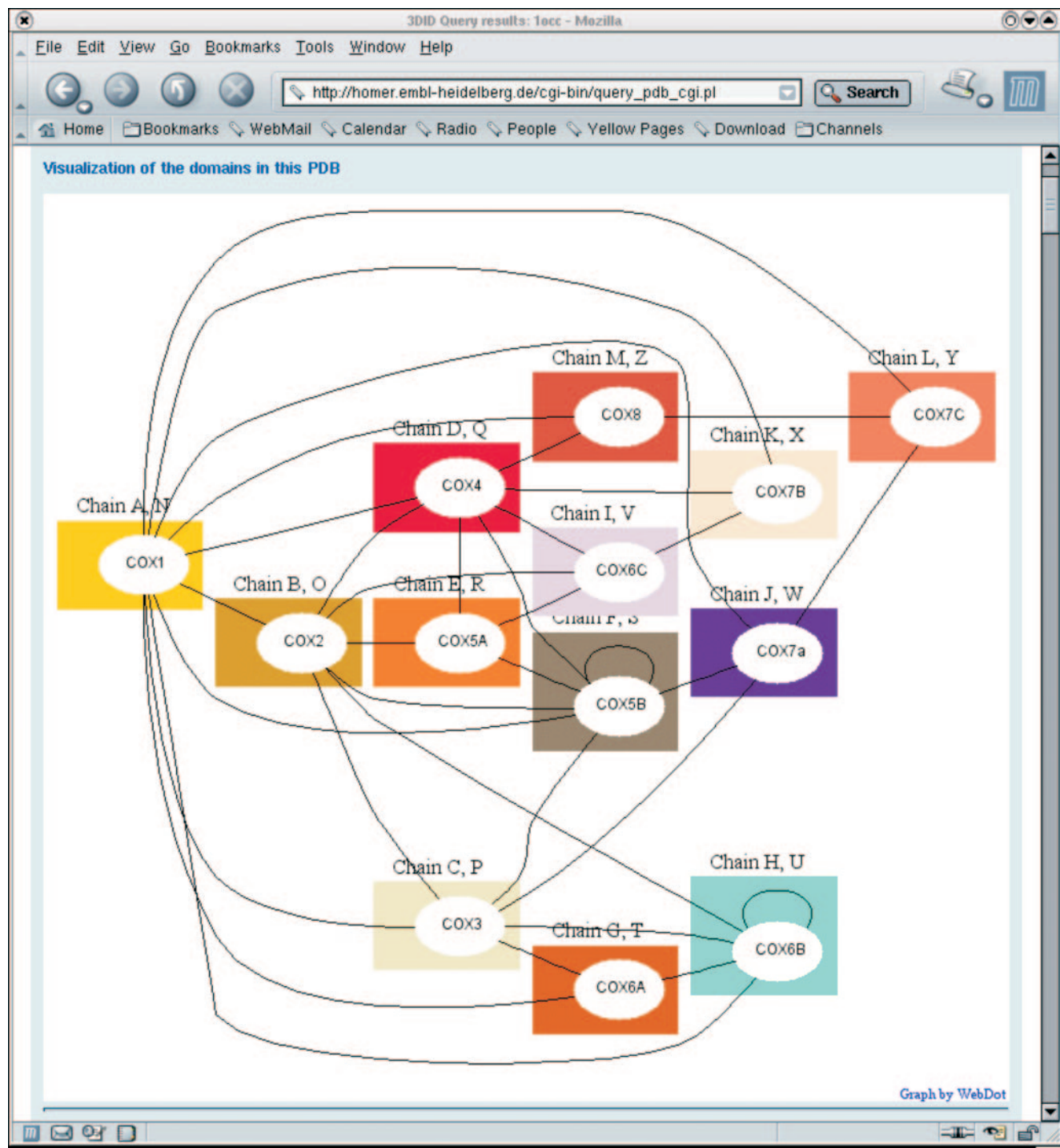
five contacts (hydrogen bonds, electrostatic or van de Waals interactions) between the two domains, and removed those that lack a significant interface as described previously (13). Nevertheless, it is likely that 3did still contains some non-biological contacts (e.g. from crystal packing), although we are working to remove them. The page will also show a list of the PDB codes for such domains and the associated functional GO terms, if defined. All the domain-domain interactions will also be displayed as an interactive network (Figure 1), where the user can choose the depth and a color scheme based on molecular function, biological process or cellular compartment as described by GO. The network also gives information on the type of contacts (i.e. intra- or inter-molecular) observed between the domains.

The user can then select a particular interaction among all the possibilities and retrieve the specific details stored in 3did. The output page for each domain-domain interaction displays a table with information concerning all the known 3D structures where this interaction is found (Figure 2). The table

shows the exact location of the two domains in the 3D complex and gives empirical potential scores and Z-scores, which provide a measure of the number of favorable interacting residue pairs at the interface (13,14). They generally account for interaction specificity: the higher the Z-score, the more specific the interaction. Finally, by clicking on the *rasmol* (15) icon, we will get a display of the 3D complex. The two interacting domains are colored and shown in *ribbons* representation with the residues participating in the interface (i.e. making hydrogen bonds, salt bridges or van der Waals contacts) are shown in *ball-and-stick* (Figure 2, top right).

The table also contains links to our tool for plotting similarity in interactions (SimInt) (16). SimInt plots structural comparisons (iRMSD) of all instances of interactions of known 3D structure, highlighting those between the domains of interest (Figure 2, bottom right). This plot provides details as to how interactions involving particular families, super-families and folds, as defined in the SCOP database (17) can vary. Based on an analysis of hundreds of interactions,





**Figure 3.** Graph of the COX complex. This shows the protein chains contained in the structure (PDB code 2occ) and the interacting domains referenced in 3did. For multiple copies of the complex, chains containing the same domains are grouped into one block. Lines connecting the domains show both inter- and intra-chain interactions as in Figure 1.

we suggested that two pairs of proteins do interact in a similar way if the iRMSD is  $<10 \text{ \AA}$ .

We have also incorporated into 3did experimental interaction data for the Yeast *Saccharomyces cerevisiae* from MIPS (Figure 2). For each yeast protein, we assign domains and whenever two interacting proteins contain domains also present in 3did, we suggest that the interaction will likely occur

via these domains, therefore suggesting molecular details for such interaction (e.g. which residues are involved, etc.). It should be noted that some interactions in MIPS (i.e. those that form pull-down experiments) link subunits in a complex that are not in physical contact and thus are not present in 3did.

The user can also choose to query 3did by pasting a protein sequence. Here, the web-tool will graphically display your

sequence with Pfam domains assigned automatically by means of BLAST (18) ( $E$ -value  $\leq 10^{-5}$ ) and links to interaction information for each domain. Alternatively, the user can search for all interactions in a given structure (Figure 3) or query 3did directly with GO or SCOP accession codes.

3did also offers the possibility to check whether there is a putative indirect interaction path across similar proteins of known structure. The search engine looks for all possible paths in 3did and displays those with the shortest length. This is particularly useful for large complexes, where components are known, but not the physical contacts. For example, in cytochrome *c* oxidase (COX), we can find a path between domains COX2 and COX8 since, although they do not interact directly, both interact with COX4 (Figure 3).

Future developments will include domain definitions from SMART (19), additional experimental interaction data and classification of interaction types (transient, tight-complexes, etc.).

## AVAILABILITY

MySQL and flat files containing the entire database are available through the website for independent studies.

## REFERENCES

1. Uetz,P., Giot,L., Cagney,G., Mansfield,T.A., Judson,R.S., Knight,J.R., Lockshon,D., Narayan,V., Srinivasan,M., Pochart,P. *et al.* (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
2. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.
3. Rain,J.C., Selig,L., De Reuse,H., Battaglia,V., Reverdy,C., Simon,S., Lenzen,G., Petel,F., Wojcik,J., Schachter,V. *et al.* (2001) The protein–protein interaction map of *Helicobacter pylori*. *Nature*, **409**, 211–215.
4. Giot,L., Bader,J.S., Brouwer,C., Chaudhuri,A., Kuang,B., Li,Y., Hao,Y.L., Ooi,C.E., Godwin,B., Vitols,E. *et al.* (2003) A protein interaction map of *Drosophila melanogaster*. *Science*, **302**, 1727–1736.
5. Li,S., Armstrong,C.M., Bertin,N., Ge,H., Milstein,S., Boxem,M., Vidalain,P.O., Han,J.D., Chesneau,A., Hao,T. *et al.* (2004) A map of the interactome network of the metazoan *C. elegans*. *Science*, **303**, 540–543.
6. Ho,Y., Gruhler,A., Heilbut,A., Bader,G.D., Moore,L., Adams,S.L., Millar,A., Taylor,P., Bennett,K., Boutilier,K. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180–183.
7. Gavin,A.C., Bosche,M., Krause,R., Grandi,P., Marzioch,M., Bauer,A., Schultz,J., Rick,J.M., Michon,A.M., Cruciat,C.M. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–147.
8. Von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, **417**, 399–403.
9. Bourne,P.E., Address,K.J., Bluhm,W.F., Chen,L., Deshpande,N., Feng,Z., Flери,W., Green,R., Merino-Ott,J.C., Townsend-Merino,W. *et al.* (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.*, **32**, D223–D225.
10. Bateman,A., Coin,L., Durbin,R., Finn,R.D., Hollich,V., Griffiths-Jones,S., Khanna,A., Marshall,M., Moxon,S., Sonnhammer,E.L. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.
11. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.*, **25**, 25–29.
12. Mewes,H.W., Amid,C., Arnold,R., Frishman,D., Guldener,U., Mannhaupt,G., Munsterkotter,M., Pagel,P., Strack,N., Stumpflen,V. *et al.* (2004) MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.*, **32**, D41–D44.
13. Aloy,P. and Russell,R.B. (2002) Interrogating protein interaction networks through structural biology. *Proc. Natl Acad. Sci. USA*, **99**, 5896–5901.
14. Aloy,P. and Russell,R.B. (2003) InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics*, **19**, 161–162.
15. Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.
16. Aloy,P., Ceulemans,H., Stark,A. and Russell,R.B. (2003) The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.*, **332**, 989–998.
17. Andreeva,A., Howorth,D., Brenner,S.E., Hubbard,T.J., Chothia,C. and Murzin,A.G. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res.*, **32**, D226–D229.
18. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
19. Letunic,I., Copley,R.R., Schmidt,S., Ciccarelli,F.D., Doerks,T., Schultz,J., Ponting,C.P. and Bork,P. (2004) SMART 4.0: towards genomic data integration. *Nucleic Acids Res.*, **32**, D142–D144.